对虾养殖溶解氧浓度组合预测模型EMD-RF-LSTM

尹　航[1,3,4,6]，李祥铜[1,3,4,6]，徐龙琴[1,3,4,6]，李景彬[2]，刘双印[1,2,3,4,5,6]，
曹　亮[1,3,4]，冯大春[1,3,4,6]，郭建军[1,3,4,6]，李利桥[2*]

（1. 仲恺农业工程学院 信息科学与技术学院，广东广州 510225；2. 石河子大学 机械电气工程学院，新疆石河子，832000；3. 仲恺农业工程学院 智慧农业创新研究院，广东广州 510225；4. 仲恺农业工程学院 广东省高校智慧农业工程技术研究中心，广东广州 510225；5. 仲恺农业工程学院 广东省水禽健康养殖重点实验室，广东广州 510225；6. 仲恺农业工程学院 广东省农产品安全大数据工程技术研究中心，广东广州 510225）

摘　要：溶解氧（DO）浓度是对虾养殖水质检测的核心指标。为提高对虾养殖溶解氧浓度的预测精度，本研究提出了一种基于经验模态分解、随机森林和长短时记忆神经网络（EMD-RF-LSTM）的对虾养殖溶解氧浓度组合预测模型。首先采用经验模态分解（EMD）对养殖水质溶解氧浓度时序数据进行多尺度特征提取，得到不同尺度下的固有模态分量（IMF）；然后分别采用长短时记忆神经网络（LSTM）和随机森林（RF）对高、低频不同尺度IMF进行建模；最后结合各分量预测结果构建叠加模型，实现对溶解氧浓度时序数据的综合预测。本研究模型在广东省湛江市南三岛对虾养殖基地展开了试验及应用，在基于真实数据集的性能测试中，经验模态分解后EMD-ELM模型与极限学习机（ELM）模型对比，平均绝对误差（MAPE）、均方根误差（RMSE）和平均绝对误差（MAE）分别降低了30.11%、29.60%和32.95%。在经验模态分解基础上用RF和LSTM对不同特征尺度的本征模态分量分别预测后叠加求和，EMD-RF-LSTM模型预测的精度指标*MAPE*、*RMSE*和*MAE*分别为0.0129、0.1156和0.0844，其中关键指标*MAPE*较EMD-ELM、EMD-RF和EMD-LSTM分别降低了84.07%、57.57%和49.81%，预测精度显著提高。结果表明，本研究针对经验模态分解后高、低频分量分别预测的策略可有效提升综合性能，表明本研究模型具有较高的预测精度，能够较准确地实现对虾养殖水体中溶解氧浓度预测。

关键词：对虾养殖；溶解氧浓度预测；经验模态分解；随机森林；长短时记忆神经网络

中图分类号：TP391　　**文献标志码：**A　　**文章编号：**202106-SA008

引用格式：尹航，李祥铜，徐龙琴，李景彬，刘双印，曹亮，冯大春，郭建军，李利桥. 对虾养殖溶解氧浓度组合预测模型EMD-RF-LSTM[J]. 智慧农业(中英文), 2021, 3(2): 115-125.

YIN Hang, LI Xiangtong, XU Longqin, LI Jingbin, LIU Shuangyin, CAO Liang, FENG Dachun, GUO Jianjun, LI Liqiao. EMD-RF-LSTM: Combination prediction model of dissolved oxygen concentration in prawn culture[J]. Smart Agriculture, 2021, 3(2): 115-125. (in Chinese with English abstract)

收稿日期：2021-06-11　修订日期：2021-06-28

基金项目：国家自然科学基金项目（61871475）；广东省自然科学基金项目（2021A1515011605）；现代农业机械兵团重点实验室开放课题资助（BTNJ2021002）；广东省科技厅重点领域研发计划项目（2020B0202080002）；北京市自然科学基金项目（4182023）；广东省科技计划项目（2019B020215003）；广州市重点研发计划项目（20210300003）；广州市创新平台建设计划项目（201905010006）；广东省农业技术研发项目（2018LM2168）

作者简介：尹 航（1978—），男，博士，副教授，研究方向为人工智能和重大装备健康管理。E-mail：736028008@qq.com。

*通讯作者：李利桥（1988—），女，博士，副教授，研究方向为智慧农业和农牧机械装备研究。电话：17590396517。E-mail：liliqiao1108@163.com。

1 引 言

溶解氧（Dissolved Oxygen，DO）浓度是养殖水质检测的核心指标，影响着水产生物的生长速度和成活率，是决定对虾品质及产量的重要因素[1,2]。目前，中国典型南美白虾的平均养殖密度达到220—300尾/m^2，养殖周期为2—3期。由于虾塘养殖密度远高于鱼塘养殖密度，接近工厂鱼养殖密度，对于溶解氧浓度的监控要求更高。构建对虾养殖环境水体溶解氧变化模型，精准预测水体溶解氧浓度变化[3,4]，对于实现对虾精细化养殖管理和调控，科学决策养殖密度和饲料配比，确保对虾在无应激环境下健康生长、提高养殖效益具有重要意义[5]。

目前在水产养殖领域已有部分团队广泛开展溶解氧浓度预测方法研究。Liu等[6]采用小波分析（Wavelet Analysis，WA）、柯西粒子群优化最小二乘支持向量回归机的溶解氧浓度预测模型，并应用于河蟹养殖DO预测；徐龙琴等[7]采用小波分析进行多尺度特征提取，通过加权最小二乘支持向量回归机对不同尺度序列分别建模，实现DO预测；Huan等[8]采用梯度增强决策树和长短时记忆网络（Long Short-Term Memory，LSTM）对水产养殖溶解氧浓度进行了预测；朱南阳等[9]优化LSTM反向传播时的损失函数，提出了提高低溶解氧含量估算精度的溶解氧预测模型（LDO-LSTM），不但可以保证整体溶氧预测精度，且能提高较低溶解氧浓度值的估算精度。

在前期研究中，有研究者认为对虾养殖水体溶解氧具有长时序、不稳定、多尺度非线性等特点[10,11]；且受多因素复杂耦合关系影响[12]；难以建立高性能泛化模型[13]。由于感知设备失能、噪声干扰和长时序数据[14]，以及监测点时空分布差异[15]等问题，需要对对虾养殖水体溶解氧时序数据进行降噪、多尺度分析、时空分类及特征提取等预处理[16]。小波分析曾被用于数据降噪和特征提取，但需预定基函数，并存在人为因素干扰[17]。经验模态分解（Empirical Mode Decomposition，EMD）可将非平稳时序数据多尺度分解成低耦合模态分量（Intrinsic mode function，IMF），能有效进行数据降噪和抗干扰预处理[18-20]。

目前，EMD在水产养殖等领域得到了应用。徐龙琴等[21]将EMD和极限学习机（Extreme Learning Machine，ELM）结合，构建了基于EMD-ELM的水温组合预测模型。施珮等[22]在徐龙琴等[21]研究结果基础上，结合改进遗传算法（Improved Genetic Algorithm，IGA）和改进极限学习机（Improved Extreme Learning Machine，SELM）构建了基于EMD-IGA-SELM的预测模型，以提高水体温度预测的精度和稳定性。杨亮等[23]提出了基于EMD-LSTM的预测模型，将氨气浓度时间序列数据进行EMD处理，生成不同时间尺度下的模态分量，然后使用LSTM对各分量分别预测，再相加以实现氨气浓度的组合模型。戴邵武等[24]提出LSTM在传统神经网络基础上增加隐藏层，有效避免了梯度消失和爆炸，具有较好的预测精度和鲁棒性；赵晓东等[25]在对基于频域分解和深度学习算法的预测模型研究中发现，LSTM在高频分量预测的效果上表现优异，而在训练样本较少的低频分量预测上效果不佳。秦喜文等[26]利用经验模态分解与随机森林构建的EMD-RF模型，在不同频度分量上获得了较高的精度和泛化性能。

由以上研究可知，EMD分解和LSTM组合模型已用于溶解氧浓度预测，但在训练样本较少的低频分量预测上效果不佳的问题有待解决，针对不同频域选择合适预测模型的组合预测方法还有待进一步研究。为了解决训练样本较少时非线性时序列数据经验模态分解后不同频域模态分量预测精度不佳的问题，本研究结合经验模态分解、随机森林和长短时记忆神经网络提出了一种基于EMD-RF-LSTM的对虾养殖溶解氧非线性组合预测模型，通过EMD将养殖溶解氧时序数据进行多尺度分解，获得不同特征尺度的本征模态分量和残余分量，结合各分量预测结果，选择RF和LSTM分别对低频分量、高频分量和残差

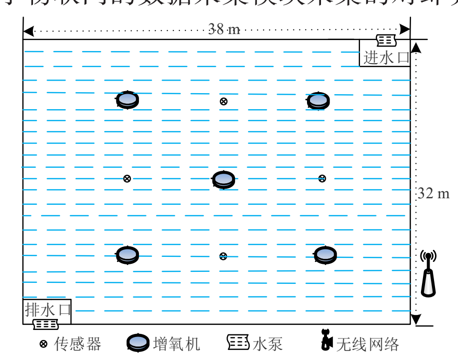
进行建模预测，最后将各预测结果叠加求和，实现对虾养殖水体溶解氧浓度预测。

2 数据与方法

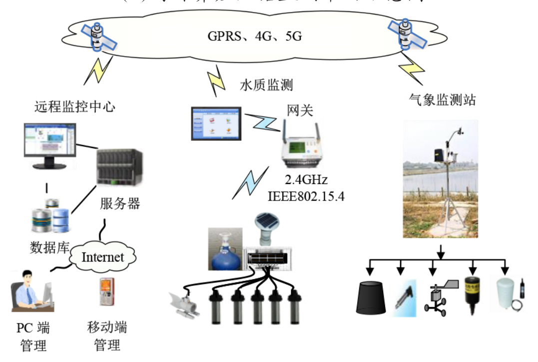
2.1 研究数据

为评估本研究模型在真实环境下的表现，本研究在广东省湛江市南三岛对虾养殖基地开展，采集对虾养殖池塘真实数据。试验用对虾养殖池塘为长 38.0 m、宽 32.0 m、水深 1.1 m，在池塘内多点部署了多参数水质传感器、增氧机、循环泵等水质监控设备。对虾养殖池塘监测平面示意图及试验平台拓扑结构图如图 1 所示。

对虾养殖环境监控及试验平台包括数据采集、无线传输、数据处理、智能监控等功能。其中基于物联网的数据采集模块采集的对虾养殖水质参数数据包括溶解氧、pH 值、水温、电导率和浊度等，采集频率为 30 min。



(a)对虾养殖池塘监测平面示意图



注：主要传感器包括水下摄像头、PH 传感器、浊度传感器、水温传感器、电导率传感器、溶解氧传感器、雨量传感器、气压传感器、风速风向、太阳辐射、温度传感器、湿度传感器

(b)试验平台拓扑结构图

图 1 对虾养殖环境物联网监控及试验平台

Fig.1 Internet of things monitoring and experimental platform for shrimp culture environment

2.2 研究方法

2.2.1 经验模态分解(EMD)

EMD 是一种自适应信号时频处理方法[18]，规避预先设置基函数等人为因素影响，能够将非平稳非线性原始信号进行自适应多尺度分解，获得一组平稳性和周期性特征的本征 IMF 和残余分量 RES[20-24]。将对虾养殖水体溶解氧原始时序数据记为 $X(t)$，则 EMD 步骤如下。

（1）通过三次样条插值法，拟合得到溶解氧原始时序数据信号的上下包络线，计算局部极大值 $X_{max}(t)$、极小值 $X_{min}(t)$ 及均值 $M(t)$，如公式（1）。

$$M(t)=\frac{X_{min}(t)+X_{max}(t)}{2} \quad (1)$$

（2）计算 $X(t)$ 与 $M(t)$ 之差 $H(t)$，如公式（2）：

$$H(t)=X(t)-M(t) \quad (2)$$

若 $H(t)$ 符合本征模态分量的要求，则增加为初始的 IMF 分量，记作 $C_1(t)$；如不符合，则作为 $X(t)$ 重复以上步骤，直至成为一个新增 IMF 分量，最终构成信号序列的高频分量；

（3）在 $H(t)$ 中减去 $C_1(t)$ 可得到残差项 r_1（t），并将其作为新的信号序列，用（2）中方法得到其余 IMF 分量 $C_2(t)$，$C_2(t)\cdots$，$C_n(t)$ 和残差项 $r_n(t)$。原始时序 $X(t)$ 最终可分解表示为各组分量和残余项 $r_n(t)$ 之和，如公式（3）。

$$x(t)=\sum_{i=1}^{n}C_i(t)+r_n(t) \quad (3)$$

2.2.2 随机森林

随机森林（Random Forest，RF）是一种继承自举集成（Bootstrap Aggregation，Bagging）算法思想的机器学习方法[27,28]，使用分类与回归决策树作为弱学习器，通过多个互相独立且权重相同的决策树组成决策森林，相比传统决策树方法，能快速收敛、有效克服过拟合，对于非线性非平稳长时序数据的预测具有较高精度，在训练速度、泛化能力和预测能力上具有优势[29]。

2.2.3 长短时记忆神经网络(LSTM)

LSTM 是由 Hochreiter 等在循环神经网络（Recurrent Neural Networks，RNN）基础上提出的神经网络模型[20,23-25]。

LSTM模型采用细胞状态和3个门结构替换RNN隐含层神经元，实现了去除或增加控制信息在细胞状态的能力，克服了RNN梯度消失、学习能力下降、信息长期依赖等问题[30,31]。

LSTM中使用输入门和遗忘门来控制单元状态向后传递的信息；输出门控制单元状态用于输出LSTM的当前值，如下：

$$\begin{cases} \boldsymbol{i}_t = \sigma\left(W_i\left[h_{t-1}, x_t\right] + b_i\right) \\ \boldsymbol{f}_t = \sigma\left(W_f\left[h_{t-1}, x_t\right] + b_f\right) \\ \boldsymbol{C}_t = \boldsymbol{f}_t C_{t-1} + i_t\left(\tanh\left(W_c\left[h_{t-1}, x_t\right] + b_c\right)\right) \\ \boldsymbol{o}_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \\ \boldsymbol{h}_t = \boldsymbol{o}_t \tanh\left(\boldsymbol{C}_t\right) \end{cases} \quad (4)$$

其中，$\boldsymbol{i}$、$\boldsymbol{f}$、$\boldsymbol{o}$、$\boldsymbol{C}$分别代表输入门、遗忘门、输出门以及候选向量；W为权重；b为偏置；σ（.）为*sigmoid*激活函数；tanh（.）为双曲正切激活函数；$\boldsymbol{i}_t$、$\boldsymbol{f}_t$、$\boldsymbol{o}_t$、$\boldsymbol{C}_t$分别代表输入门、遗忘门、输出门以及t时刻的候选向量更新值；W_c和b_c代表候选向量C的权重和偏置；x_t为t时刻序列输入，h_t为t时刻的输出。

3 基于EMD-RF-LSTM的组合预测模型设计

3.1 模型设计

为解决非线性时序列数据经验模态分解后不同频域模态分量预测精度不佳的问题，验证按高、低频分量分别预测的效果，本研究设计了基于EMD-RF-LSTM的对虾养殖水体溶解氧组合预测模型，并选用溶解氧浓度数据作为输入。首先采用EMD对呈现周期波动的、非线性的对虾养殖溶解氧时间序列数据进行多尺度分解，划分成高频IMF、低频IMF及残差值RES；然后对分解后数据进行归一化处理，划分训练集和测试集；使用低频分量训练RF模型，高频分量训练LSTM模型，并用Adam反复优化调整LSTM模型参数；最后，将测试集用于该模型评估，并展开与ELM、RF、LSTM等标准模型及采用EMD分解模型的对比试验，以验证本研究模型对对虾养殖水体溶解氧的预测性能。详细步骤如下。

（1）通过水质检测传感器采集溶解氧时间序列数据，完成预处理；

（2）对预处理后溶解氧时序数据进行EMD分解，得到不同频率IMF分量，并归一化处理；

（3）将归一化处理后的对虾养殖溶解氧IMF分量分为高频和低频，并划分训练集和测试集；

（4）对IMF高频分量、低频分量及余量，分别建立LSTM及RF模型，对预测模型参数和权重进行初始化；

（5）将训练集作为输入对模型进行训练，对LSTM模型参数及权重进行迭代优化处理，完成基于EMD-RF-LSTM的对虾养殖溶解氧预测模型构建；

（6）测试预测模型，并与其它模型对比。

所构建的预测模型如图2所示。

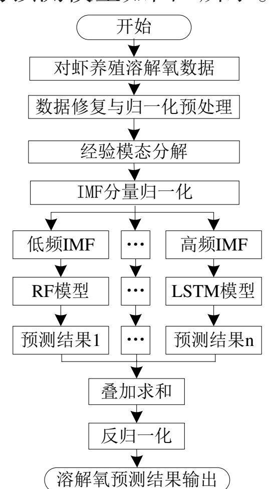


图2 基于EMD-RF-LSTM的对虾养殖溶解氧预测模型

Fig. 2 Flowchart of dissolved oxygen prediction model based on EMD-RF-LSTM

3.2 评价指标

为验证EMD-RF-LSTM模型对对虾养殖水体溶解氧浓度的预测性能，展开了本模型与其它模

型的对比试验。选择了平均绝对百分比误差（MAPE）、均方根误差（RMSE）和平均绝对误差（MAE）三项评价指标对组合模型的预测性能进行性能评价，并开展对比。

4 试验及结果分析

4.1 数据预处理

试验以湛江市南三岛对虾养殖基地试验池塘溶解氧浓度为研究对象，以基于物联网数据采集模块采集的对虾养殖水质数据作为试验样本。选取2020年7月20日至8月20日采集的共计1488个样本加入试验用数据集，并取前1344条数据作为训练集，最后144条数据作为测试集。图3显示了完整数据采集周期的原始数据，其中横坐标为每30 min间隔采集的数据序列，纵坐标为溶解氧浓度数值。由图3可见，真实对虾养殖现场水体溶解氧浓度时序数据呈现显著周期性、非线性特征。

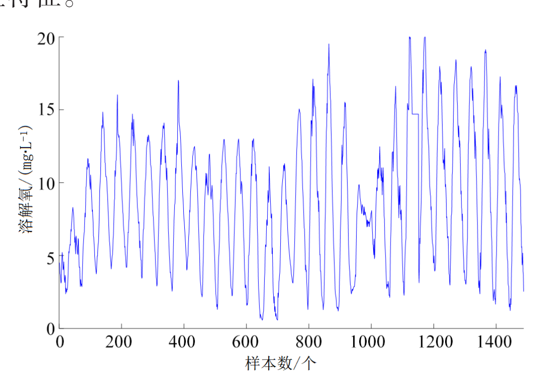


图3 溶解氧原始数据变化曲线图

Fig.3 Variation curve of the original air oxygen data

针对水质传感器故障等因素导致的采集数据异常，利用均值平滑法进行处理。如果存在参数与其平均值之差的绝对值大于其标准差的3倍，即断定为异常值并用其两侧数据的平均值替换，如公式（5）所示。

$$\begin{cases} \left| P_t - \dot{P} \right| > 3\delta_{P_t} \\ P_t' = \dfrac{P_{t-1} + P_{t+1}}{2} \end{cases} \tag{5}$$

其中，P_t为t时刻溶解氧参数采集值；P_t'为异常数据处理后值；P为水体溶解氧数据序列均值。

为提高预测准确率减少误差，便于研究对虾养殖溶解氧浓度数据间的相关性，更好地提取时序数据信息，本研究利用公式（6）对数据进行归一化处理。

$$N'' = \frac{N - N_{\min}}{N_{\max} - N_{\min}} \tag{6}$$

其中，N_{max}为溶解氧浓度最大值，$N_{\min}$为最小值，单位mg/L；N'为归一化值。

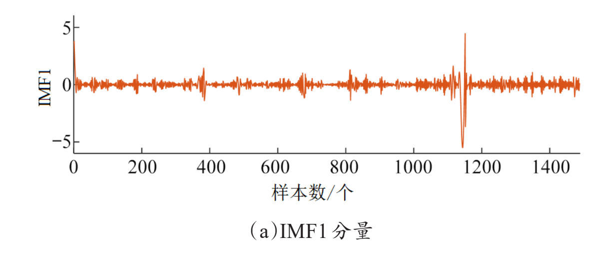
4.2 开发环境及工具选择

试验计算机环境为Intel I7-7700K CPU，8GB内存，Window7 + python3.7 + MATLAB，集成开发环境为Anaconda3。

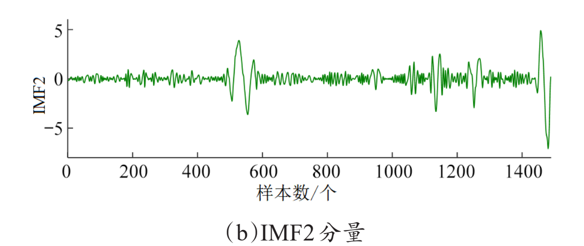
其中，EMD和ELM模型基于MATLAB工具箱实现，RF模型基于Anaconda的Sklearn程序包实现，LSTM模型基于Keras框架构建，试验参数采用留一法交叉验证网格搜索法（Leave-One-Out-Cross-Validation，LOOCV）优化。

4.3 基于EMD的溶解氧多尺度分解

为得到更加准确的预测效果，获得高精度的对虾养殖溶解氧时序分量，本研究首先使用EMD对原始溶解氧时序数据进行多尺度分解，分解后得到的分量如图4所示。



（a）IMF1分量



（b）IMF2分量

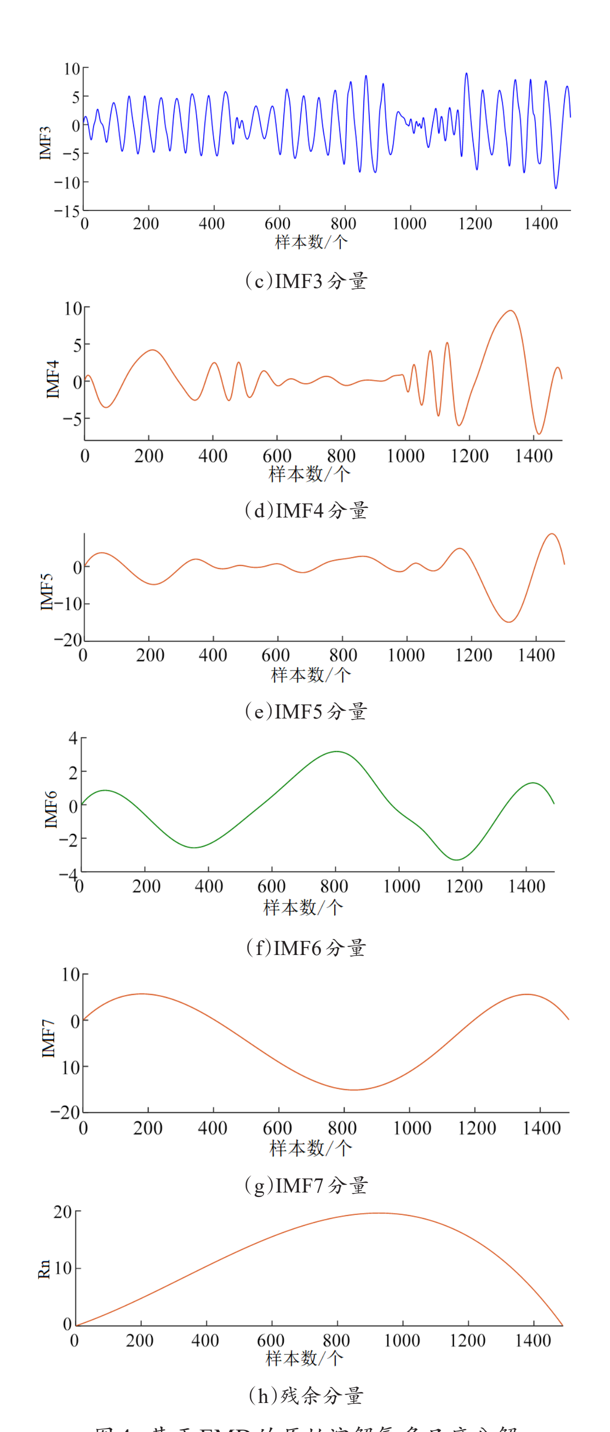


图 4　基于EMD的原始溶解氧多尺度分解

Fig. 4　Multi-scale decomposition of original dissolved oxygen based on EMD

由图 4 可以看出，对虾养殖水体溶解氧浓度时序数据不同尺度的特征明显，分解后得到的本征模态分量IMF1—IMF7各表现出不同的信息特征，最后的剩余分量序列平稳，体现出对虾养殖水体溶解氧总体含量的长期变化状态。

4.4 IMF 分量预测及参数设置

基于文献［25］、［30］和［31］，本研究在经验模态分解的基础上，采用全面试验方法利用LSTM模型和RF模型分别对IMF1—IMF7及RES分量进行建模训练，并使用扩展随机梯度下降法对LSTM模型参数进行优化，利用网格搜索法对RF模型参数进行优化，以寻找LSTM和RF模型在不同频度分量上的预测表现。

标准LSTM隐含层节点数、批尺度和时间步分别为20、32和5，RF模型的参数学习率为0.1，节点数深度为3，节点数颗数为500，最小叶子权重为6。得到各个IMF分量和RES分量的LSTM模型和RF模型预测结果分别如表1和表2所示。

表 1　各分量LSTM预测结果精度分析

Table 1　Forecast results for each sub sequence of LSTM

分量	*MAPE*	*RMSE*	*MAE*
IMF1	0.8097	0.2911	0.2350
IMF2	0.1397	0.4436	0.1687
IMF3	0.0100	0.0543	0.0449
IMF4	0.0082	0.0880	0.0435
IMF5	0.0185	0.1298	0.1424
IMF6	0.0206	0.0176	0.0170
IMF7	0.0033	0.0080	0.0080
Rn	0.0101	0.0353	0.0331

注：Rn为残余分量

表 2　各分量RF预测结果精度分析

Table 2　Forecast results for each sub sequence of RF

分量	*MAPE*	*RMSE*	*MAE*
IMF1	1.1542	0.3802	0.3268
IMF2	0.1218	0.0952	0.0604
IMF3	0.0118	0.0520	0.0421
IMF4	0.0154	0.0796	0.0524
IMF5	0.0044	0.0234	0.0194
IMF6	0.0007	0.0004	0.0003
IMF7	0.0045	0.0099	0.0084
Rn	0.0097	0.0253	0.0249

由表1和表2可见，RF模型对高频分量IMF1的*MAPE*值仅为1.1542，对IMF4的*MAPE*值为0.0154，均低于LSTM对应分量的预测精度。但随着各分量频率降低，RF模型的预测精度也随之提高；此时随着分量频率降低，LSTM

模型预测精度呈现下降趋势。两者相对比发现，在高频分量IMF1—IMF4的预测精度上，LSTM模型在关键指标上优于RF模型，而RF模型则在低频分量上表现更好，试验结果符合预期。由此结果可知，低频分量适合训练RF模型，高频分量适合训练LSTM模型。

4.5 基于EMD-RF-LSTM的组合预测

依据4.4节试验展现特点，采用LSTM和RF模型分别对高频分量（IMF1—IMF4）、低频分量和残差（IMF5—IMF7，Rn）进行建模，然后将各分量预测结果求和，以实现基于EMD-RF-LSTM的对虾养殖溶解氧浓度预测。

为验证模型性能，分别采用标准模型、模态分解后模型以及本研究模型，使用相同数据集开展溶解氧浓度预测。其中标准ELM模型采用sigmoid激活函数，隐含层节点数为8。不同模型的溶解氧浓度预测结果如图5所示，各指标如表3所示。

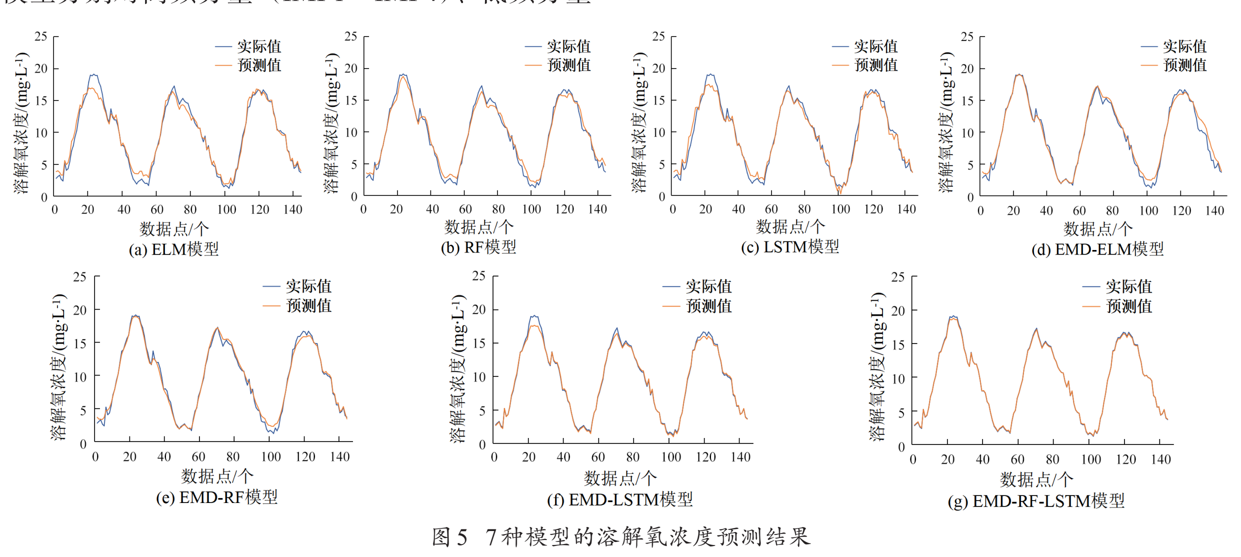


图5 7种模型的溶解氧浓度预测结果

Fig.5 Oxygen forecasting results of the seven models

表3 7种模型预测结果精度

Table 3 Forecast results for the seven models

序号	模型	*MAPE*	*RMSE*	*MAE*
1	ELM	0.1159	0.6969	0.5824
2	RF	0.1027	0.7585	0.5823
3	LSTM	0.1021	0.6794	0.5070
4	EMD-ELM	0.0810	0.4906	0.3905
5	EMD-RF	0.0304	0.3803	0.2467
6	EMD-LSTM	0.0257	0.3173	0.2113
7	**EMD-RF-LSTM**	**0.0129**	**0.1156**	**0.0844**

4.6 结果分析

4.6.1 经验模态分解(EMD)分析

对试验结果进行分析统计，在相同对虾养殖溶解氧浓度数据集下：EMD-ELM模型与标准ELM模型对比，*MAPE*、*RMSE*和*MAE*指标分别降低了30.11%、29.60%和32.95%；EMD-RF与标准RF模型对比，*MAPE*、*RMSE*和*MAE*指标分别降低了70.40%、49.86%和57.63%；EMD-LSTM与标准LSTM对比，*MAPE*、*RMSE*和*MAE*指标分别降低了74.83%、53.30%和58.32%。

以关键精度指标*MAPE*为例，采用EMD分解ELM、RF和LSTM模型相比对应标准模型分别降低了30.11%、70.40%和74.83%，预测精度显著高于标准模型，证明基于EMD的时序数据多尺度分解可有效提升预测性能。

4.6.2 多频度模态分量组合预测分析

由试验结果可知，在相同数据集下，在经验模态分解（EMD）基础上通过RF和LSTM对不同特征尺度的本征模态分量分别预测的EMD-

RF-LSTM模型与4.6.1中EMD分解后的各模型对比。以关键精度指标*MAPE*为例，本研究提出的基于多频模态分量组合预测模型，较普通EMD分解后模型分别降低了84.07%、57.57%和49.81%，预测精度显著提高，证明针对多频分量的预测策略可提升模型性能。

4.6.3 基于EMD-RF-LSTM组合预测模型分析

对预测结果分析可发现，经验模态分解具备多尺度提取对虾养殖溶解氧时间序列信息的特性，数据分解之后会在保留原始信息基础上得到更多的本征模态系数时间序列信号；而RF可有效提取低频IMF数据信息，LSTM模型对高频数据有理想的效果，对于时间序列信息能够高效利用。

本研究提出的基于EMD-RF-LSTM组合模型结合了经验模态分解的多尺度特征提取、LSTM对长时间序列高频数据预测以及RF算法对低频IMF数据信息提取的优势，能获得较高的对虾养殖水体溶解氧浓度预测精度，预测曲线能够很好地拟合养殖溶解氧浓度非线性时间序列变化趋势，取得很好的预测效果。

5 讨论与结论

5.1 讨论

在对对虾养殖水质长期检测数据的观察中发现，养殖水质尤其是溶解氧（DO）浓度变化相对缓慢，在文献［7］和［12］中均可见溶解氧浓度在30 min内变化很小。相对于文献［30］中数据采集周期、采集间隔及训练数据量，本研究在检测周期同为一个月的情况下，适当增大了数据采集间隔，以减少用于训练的总样本数量。由于LSTM作为一种时间循环神经网络是为解决RNN存在长期依赖问题而设计，对于时间序列数据有较好的记忆能力，对于长度较短的时间序列数据也具有一定预测效果。在文献［25］中，LSTM模型在经验模式分解后的高频分量上预测效果表现优异，而在训练样本较少的低频分量上预测效果不佳；而文献［31］为验证训练样本较少的训练效果，为提出的EMD-LSTM模型选择了1500组数据作为训练样本，并获得较好预测效果。综上，本研究综合考虑变量数量、总体样本量比例关系，为验证本研究提出EMD-RF-LSTM模型在训练样本较少情况下的表现，从现场数据中选择了采集周期为一个月、采样间隔为30 min、共计1488组溶解氧浓度时序数据作为训练样本开展研究。

本研究在选择训练样本较少的情况下，首先验证LSTM在经验模式分解后的低频分量上预测效果不佳的情况，然后通过试验结果将IMF1—IMF4划分为适合LSTM模型训练的高频分量，将IMF5—IMF7及Rn划分为适合RF模型训练的分量，并构建了EMD-RF-LSTM组合模型以提升预测精度。此外，本研究利用历史数据进行交叉验证，模型展现了较好预测结果，为进一步验证在训练样本较少时历史数据的影响，在后续试验中将加入实际现场测试结果对本模型性能进行验证；并调整可能会引起溶氧剧烈变化的时刻的采样频率，如投饵时，或早晚，或天气变化时，调整采样间隔。

5.2 结论

本研究针对对虾养殖水体溶解氧浓度采集数据不稳定和多尺度特征等特点，分析了训练样本较少情况下非线性时序列数据经验模态分解后不同频域模态分量预测精度不佳的问题，利用EMD对对虾水质溶解氧浓度数据进行多尺度分解，使用LSTM用于高频分量预测、RF用于低频分量预测，对不同频段数据分量进行分别建模预测，通过真实养殖环境数据试验证明，本研究提出基于EMD-RF-LSTM的组合预测模型的*MAPE*、*RMSE*和*MAE*指标分别为0.0129、0.1156和0.0844，与经验模式分解后的EMD-ELM、EMD-RF和EMD-LSTM模型相比关键指标*MAPE*分别降低了84.07%、57.57%和49.81%，在训练样本较少的情况下对于对虾养殖水体溶解

氧浓度具有良好的预测效果，有效提高了预测精度和鲁棒性。

参考文献：

[1] 于道德, 宋静静, 叶海斌, 等. 生态理念下对虾健康养殖发展建议[J]. 生态学杂志, 2021, 40(2): 568-576.
YU D, SONG J, YE H, et al. Suggestions on the development of healthy shrimp culture under ecological framework[J]. Chinese Journal of Ecology, 2021, 40 (2): 568-576.

[2] 朱林, 车轩, 刘兴国, 等. 简易式工厂化循环水对虾养殖系统构建及试验[J]. 农业工程学报, 2020, 36 (15): 210-216.
ZHU L, CHE X, LIU X, et al. Construction and experiment of simple industrial recirculating water shrimp culture system[J]. Transactions of the CSAE, 2020, 36 (15): 210 -216.

[3] 吴静, 李振波, 朱玲, 等. 融合ARIMA模型和GAWNN的溶解氧含量预测方法[J]. 农业机械学报, 2017, 48 (S1): 205-210.
WU J, LI Z, ZHU L, et al. Hybrid model of ARIMA model and GAWNN for dissolved oxygen content prediction[J]. Transactions of the CSAM, 2017, 48(S1): 205-210.

[4] 陈英义, 程倩倩, 方晓敏, 等. 主成分分析和长短时记忆神经网络预测水产养殖水体溶解氧[J]. 农业工程学报, 2018, 34(17): 183-191.
CHEN Y, CHENG Q, FANG X, et al. Principal component analysis and long short-term memory neural network for predicting dissolved oxygen in water for aquaculture[J]. Transactions of the CSAE, 2018, 34 (17): 183-191.

[5] SWETA B, AMBIKA H, MAHESHAWARI J., Simple rapid and on spot dye-based sensor for the detection of Vibrio load in shrimp culture farms[J]. Archives of Clinical Microbiology, 2021 (1): 1-8.

[6] LIU S, XU L, JIANG Y, et al. A hybrid WA-CPSO-LSSVR model for dissolved oxygen content prediction in crab culture[J]. Engineering Applications of Artificial Intelligence, 2014, 29(3): 114-124.

[7] 徐龙琴, 陈跃霞, 张军, 等. 基于WA-ABC-WLSSVR的南美白对虾工厂化育苗溶解氧预测模型[J]. 武汉大学学报(工学版), 2017, 50(4): 608-617.
XU L, CHEN Y, ZHANG J, et al. Prediction model of dissolved oxygen in industrialized vannamei breeding based on wavelet analysis and weighted least squares support vector regression optimized by artificial bee colony algorithm[J]. Engineering Journal of Wuhan University, 2017, 50(4): 608-617.

[8] HUAN J, LI H, LI M, et al. Prediction of dissolved oxygen in aquaculture based on gradient boosting decision tree and long short-term memory network: A study of Changzhou fishery demonstration base, China[J]. Computers and Electronics in Agriculture, 2020, DOI: 10.1016/j.compag.2020.105530.

[9] 朱南阳, 吴昊, 尹达恒, 等. 基于长短时记忆网络(LSTM)的蟹塘溶解氧估算优化方法[J]. 智慧农业(中英文), 2019, 1(3): 67-76.
ZHU N, WU H, YIN D, et al. An improved method for estimating dissolved oxygen in crab ponds based on Long Short-Term Memory[J]. Smart Agriculture, 2019, 1(3): 67-76.

[10] CAO W, HUAN J, LIU C, et al. A combined model of dissolved oxygen prediction in the pond based on multiple-factor analysis and multi-scale feature extraction[J]. Aquacultural Engineering, 2019, 84: 50-59.

[11] RAHMAN A, DABROWSKI J, MCCULLOCH J. Dissolved oxygen prediction in prawn ponds from a group of one step predictors[J]. Information Processing in Agriculture, 2020, 7(2): 307-317.

[12] 施珮, 匡亮, 袁永明, 等. 基于改进极限学习机的水体溶解氧预测方法[J]. 农业工程学报, 2020, 36 (19): 225-232.
SHI P, KUANG L, YUAN Y, et al. Dissolved oxygen prediction for water quality of aquaculture using improved ELM network[J]. Transactions of the CSAE, 2020, 36(19): 225-232.

[13] REN Q, WANG X, LI W, et al. Research of dissolved oxygen prediction in recirculating aquaculture systems based on deep belief network[J]. Aquacultural Engineering, 2020, 90: ID 102085.

[14] LIU Y, ZHANG Q, SONG L, et al. Attention-based recurrent neural networks for accurate short-term and long-term dissolved oxygen prediction[J]. Computers and Electronics in Agriculture, 2019, 165: ID 104964.

[15] 樊宇星, 任妮, 田港陆, 等. 基于DeepAR-RELM的池塘溶解氧时空预测方法研究[J]. 农业机械学报, 2020, 51(S1): 405-412.
FAN Y, REN N, TIAN G, et al. Spatio-temporal prediction method of dissolved oxygen in ponds based on DeepAR-RELM[J]. Transactions of the CSAM, 2020, 51(S1): 405-412.

[16] 饶伟, 杨卫中, 位耀光, 等. 鱼菜共生水体溶解氧时空变化规律及其影响因素研究[J]. 农业机械学报, 2017, 48(S1): 374-380.
RAO W, YANG W, WEI Y, et al. Temporal and spatial variability of water dissolved oxygen with influence

factors in aquaponics system[J]. Transactions of the CSAM, 2017, 48(S1): 374-380.

[17] 于辉辉. 基于机器学习的池塘养殖水质关键因子预测方法研究[D]. 北京: 中国农业大学, 2017.
YU H. Research on prediction method of key factors of pond aquaculture water quality based on machine learning[D]. Beijing: China Agricultural University, 2017.

[18] 胡茑庆, 陈徽鹏, 程哲, 等. 基于经验模态分解和深度卷积神经网络的行星齿轮箱故障诊断方法[J]. 机械工程学报, 2019, 55(7): 9-18.
HU N, CHEN H, CHENG Z, et al. Fault diagnosis for planetary gearbox based on EMD and deep convolutional neural networks[J]. Journal of Mechanical Engineering, 2019, 55(7): 9-18.

[19] 杨建华, 韩帅, 张帅, 等. 强噪声背景下滚动轴承微弱故障特征信号的经验模态分解[J]. 振动工程学报, 2020, 33(3): 582-589.
YANG J, HAN S, ZHANG S, et al. Fault diagnosis for planetary gearbox based on EMD and deep convolutional neural networks[J]. Journal of Vibration Engineering, 2020, 33(3): 582-589.

[20] 刘云, 许自强, 董王英, 等. 基于经验模态分解和长短期记忆神经网络的变压器油中溶解气体浓度预测方法[J]. 中国电机工程学报, 2019, 39(13): 3998-4008.
LIU Y, XU Z, DONG W, et al. Concentration prediction of dissolved gases in transformer oil based on empirical mode decomposition and long short-term memory neural networks[J]. Proceedings of The Chinese Society for Electrical Engineering, 2019, 39(13): 3998-4008.

[21] 徐龙琴, 张军, 李乾川, 等. 基于EMD和ELM的工厂化育苗水温组合预测模型[J]. 农业机械学报, 2016, 47(4): 265-271.
XU L, ZHANG J, LI Q, et al. Combined prediction model of water temperature in industrialized cultivation based on empirical mode decomposition and extreme learning machine[J]. Transactions of the CSAM, 2016, 47(4): 265-271.

[22] 施珮, 袁永明, 匡亮, 等. 基于EMD-IGA-SELM的池塘养殖水温预测方法[J]. 农业机械学报, 2018, 49(11): 312-319.
SI P, YUAN Y, KUANG L, et al. Water temperature prediction in pond aquaculture based on EMD-IGA-SELM neural network[J]. Transactions of the CSAM, 2018, 49(11): 312-319.

[23] 杨亮, 刘春红, 郭昱辰, 等. 基于EMD-LSTM的猪舍氨气浓度预测研究[J]. 农业机械学报, 2019, 50 (S1): 353-360.
YANG L, LIU C, GUO Y, et al. Prediction of ammonia concentration in fattening piggery based on EMD-LSTM[J]. Transactions of the CSAM, 2019, 50(S1): 353-360.

[24] 戴邵武, 陈强强, 刘志豪, 等. 基于EMD-LSTM的时间序列预测方法[J]. 深圳大学学报(理工版), 2020, 37 (3): 265-270.
DAI S, CHEN Q, LIU Z, et al. Time series prediction based on EMD-LSTM model[J]. Journal of Shenzhen University (Science & Engineering), 2020, 37(3): 265-270.

[25] 赵晓东, 苏公瑾, 李克利, 等. 一种融合EMD分解和LSTM网络的频谱占用度预测模型[J]. 计算机科学, 2020, 47(S1): 294-298.
ZHAO X, SU G, LI K, et al. Spectrum occupancy prediction model based on EMD decomposition and LSTM networks[J]. Computer Science, 2020, 47(S1): 294-298.

[26] 秦喜文, 吕思奇, 李巧玲, 等. 利用整体经验模态分解和随机森林的脑电信号分类研究[J]. 中国生物医学工程学报, 2018, 37(6): 665-672.
QIN X, LYU S, LI Q. Recognition of EEG based on ensemble empirical mode decomposition and random forest[J]. Chinese Journal of Biomedical Engineering, 2018, 37(6): 665-672.

[27] ANNA C, SALAH S, BRETT W. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review[J]. Computers and Electronics in Agriculture, 2018, 151: 61-69.

[28] DHIVYA E, DURAI R, VISHAL S, et al. Forecasting yield by integrating agrarian factors and machine learning models: A survey[J]. Computers and Electronics in Agriculture, 2018, 155: 257-282.

[29] 刘峻明, 和晓彤, 王鹏新, 等. 长时间序列气象数据结合随机森林法早期预测冬小麦产量[J]. 农业工程学报, 2019, 35(6): 158-166.
LIU J, HE X, WANG P, et al. Early prediction of winter wheat yield with long time series meteorological data and random forest method[J]. Transactions of the CSAE, 2019, 35(6): 158-166.

[30] 陈英义, 方晓敏, 梅思远, 等. 基于WT-CNN-LSTM的溶解氧含量预测模型[J]. 农业机械学报, 2020, 51 (10): 284-291.
CHEN Y, FANG X, MEI S, et al. Principal component analysis and long short-term memory neural network for predicting dissolved oxygen in water for aquaculture[J]. Transactions of the CSAM, 2020, 51(10): 284-291.

[31] 金秀章, 刘岳, 于静, 等. 基于变量选择和EMD-

LSTM网络的出口SO_2浓度预测[J/OL]. 中国电机工程学报, (2021-04-19). http://kns.cnki.net/ kcms/detail/11.2107.TM.20210416.1541.009.html.
JIN X, LIU Y, YU J, et al. Prediction of outlet SO_2 concentration based on variable selection and EMD-LSTM[J/OL]. Proceedings of The Chinese Society for Electrical Engineering, (2021-04-19). http://kns. cnki.net/kcms/detail/11.2107.TM.20210416.1541.009.html.

EMD-RF-LSTM: Combination Prediction Model of Dissolved Oxygen Concentration in Prawn Culture

YIN Hang[1,3,4], LI Xiangtong[1,3,4], XU Longqin[1,3,4], LI Jingbin[2], LIU Shuangyin[1,2,3,4,5], CAO Liang[1,3,4], FENG Dachun[1,3,4], GUO Jianjun[1,3,4], LI Liqiao[2*]

（1. *Zhongkai University of Agriculture and Engineering, College of Information Science and Technology, Guangzhou,* 510225, *China*; 2. *Shihezi University, College of Mechanical and Electric Engineerings, Shihezi,* 832000, *China*; 3. *Zhongkai University of Agriculture and Engineering, Academy of Smart Agricultural Engineering Innovations, Guangzhou* 510225, *China*; 4. *Zhongkai University of Agriculture and Engineering, Smart Agriculture Engineering Research Center of Guangdong Higher Education Institutes, Guangzhou* 510225, *China*; 5. *Zhongkai University of Agriculture and Engineering, Guangdong Key Laboratory of Waterflow Health Breeding, Guangzhou* 510225, *China*）

Abstract: Dissolved oxygen is an important environmental factor for prawn breeding. In order to improve the prediction accuracy of dissolved oxygen concentration in prawn pond, and solve the problem of low prediction accuracy of different frequency domain modal classification after empirical modal decomposition of nonlinear time series data when there are few training samples, an combination prediction model based on empirical mode decomposition (EMD), random forest (RF) and long short term memory neural network (LSTM) was proposed in this research. Firstly, the time series data of prawn breeding dissolved oxygen concentration were decomposed at multiple scales by EMD to obtain a set of stationary intrinsic mode function (IMF). Secondly, with fewer training samples, poor predicts effects on the low-frequency were verified component by LSTM. Then, IMF1—IMF4 were divided into high-frequency components through test results and used for LSTM model. IMF5—IMF7, Rn were divided for RF model, the EMD-RF-LSTM combination model was constructed to improve the prediction accuracy. Modeled low-frequency and high-frequency components IMF using RF and LSTM, then predictions of each component were accumulated and the prediction value of dissolved oxygen of sequence data were got. Finally, the performance of the model was compared with the limit learning machine (ELM), RF, standard LSTM, EMD-ELM and EMD-RF, EMD-LSTM, etc. In the test based on real dataset, the EMD-ELM model contrasted with ELM model, reduced the mean absolute error (MAPE), root mean square error (RMSE) and mean absolute error (MAE) by 30.11%, 29.60% and 32.95%, respectively. The *MAPE*, *RMSE*, *MAE* for the proposed models were 0.0129, 0.1156, 0.0844, respectively. MAPE decreased by 84.07%, 57.57%, and 49.81% compared with EMD-ELM, EMD-RF and EMD-LSTM, respectively, the prediction accuracy was significantly improved. The results show that the proposed model EMD-RF-LSTM has good prediction performance and generalization ability, which is meets the actual demand of accurate prediction of dissolved oxygen concentration in prawn culture, and can provide reference for the prediction and early warning of prawn pond water quality.

Key words: prawn pond; dissolved oxygen prediction; empirical mode decomposition; random forest; short and long-term memory neural network